

On the representation by bivariate ridge functions

Rashid A. Aliev

Institute of Mathematics and Mechanics, NAS of Azerbaijan, Baku, Azerbaijan
Baku State University, Baku, Azerbaijan
e-mail: aliyevrashid@hotmail.ru

Vugar E. Ismailov*

Institute of Mathematics and Mechanics, NAS of Azerbaijan, Baku, Azerbaijan
Baku Business University, Baku, Azerbaijan
e-mail: vugaris@mail.ru

Abstract. We consider the problem of representation of a bivariate function by sums of ridge functions. We show that if a function of a certain smoothness class is represented by a sum of finitely many, arbitrarily behaved ridge functions, then it can also be represented by a sum of ridge functions of the same smoothness class. As an example, this result is applied to a homogeneous constant coefficient partial differential equation.

2010 Mathematics Subject Classification: 26B40, 35C99, 35L25.

Keywords: Cauchy functional equation; ridge function; plane wave; representation; smoothness.

1. Introduction

Last 30 years have seen a growing interest in the study of special multivariate functions called ridge functions. This interest is due to applicability of such functions in various research areas. A *ridge function* is a multivariate function of the form

$$g(\mathbf{a} \cdot \mathbf{x}) = g(a_1x_1 + \cdots + a_nx_n),$$

where $g : \mathbb{R} \rightarrow \mathbb{R}$ and $\mathbf{a} = (a_1, \dots, a_n)$ is a fixed vector (direction) in $\mathbb{R}^n \setminus \{\mathbf{0}\}$. These functions and their linear combinations find applications in computerized tomography (see, e.g., [10, 13, 16]), in statistics (especially, in the theory of projection pursuit and projection regression; see, e.g., [3, 5]) and in the theory of neural networks (see, e.g., [8, 15, 20]). Ridge functions are also widely used in modern approximation theory as an effective and convenient tool for approximating complicated multivariate functions (see, e.g., [6, 12, 14, 17]). We refer the reader to Pinkus [18] for more on ridge functions and application areas.

It should be remarked that ridge functions have been used in the theory of partial differential equations under the name of *plane waves* (see, e.g., [9]). In general, linear combinations of ridge functions with fixed directions occur in the study of hyperbolic constant coefficient partial differential equations. For example, assume that (α_i, β_i) , $i =$

*Corresponding author

$1, \dots, r$, are pairwise linearly independent vectors in \mathbb{R}^2 . Then the general solution to the homogeneous equation

$$\prod_{i=1}^r \left(\alpha_i \frac{\partial}{\partial x} + \beta_i \frac{\partial}{\partial y} \right) u(x, y) = 0 \quad (1.1)$$

are all functions of the form

$$u(x, y) = \sum_{i=1}^r v_i(\beta_i x - \alpha_i y) \quad (1.2)$$

for arbitrary univariate functions v_i , $i = 1, \dots, r$, from the class $C^r(\mathbb{R})$.

Note that the solution of Eq. (1.1) is the sum of bivariate ridge functions. Sums of bivariate ridge functions also occur in basic mathematical problems of computerized tomography. For example, Logan and Shepp [9] (the term “ridge function” was coined by them) considered the problem of reconstructing a given but unknown function $f(x, y)$ from its integrals along certain lines in the plane. More precisely, let D be the unit disk in the plane and a function $f(x, y)$ be square integrable and supported on D . We are given projections $P_f(t, \theta)$ (integrals of f along the lines $x \cos \theta + y \sin \theta = t$) and looking for a function $g = g(x, y)$ of minimum L_2 norm, which has the same projections as f : $P_g(t, \theta_j) = P_f(t, \theta_j)$, $j = 0, 1, \dots, n-1$, where angles θ_j generate equally spaced directions, i.e. $\theta_j = \frac{j\pi}{n}$, $j = 0, 1, \dots, n-1$. The authors of [9] showed that this problem of tomography is equivalent to the problem of L_2 -approximation of the function f by sums of bivariate ridge functions with equally spaced directions $(\cos \theta_j, \sin \theta_j)$, $j = 0, 1, \dots, n-1$. They gave a closed-form expression for the unique function $g(x, y)$ and showed that the unique polynomial $P(x, y)$ of degree $n-1$ which best approximates f in $L_2(D)$ is determined from the above n projections of f and can be represented as a sum of n bivariate ridge functions.

In this paper, we are interested in the problem of smoothness in representation by sums of bivariate ridge functions with finitely many fixed directions. Assume we are given n pairwise linearly independent directions (a_i, b_i) , $i = 1, \dots, n$, in \mathbb{R}^2 and a function $F : \mathbb{R}^2 \rightarrow \mathbb{R}$ of the form

$$F(x, y) = \sum_{i=1}^n g_i(a_i x + b_i y). \quad (1.3)$$

Assume in addition that F is of a certain smoothness class, what can we say about the smoothness of g_i ? The case $n = 1$ is obvious. In this case, if $F \in C^k(\mathbb{R}^2)$, then for a vector $(c, d) \in \mathbb{R}^2$ satisfying $a_1 c + b_1 d = 1$ we have that $g_1(t) = F(ct, dt)$ is in $C^k(\mathbb{R})$. The same argument can be carried out for the case $n = 2$. In this case, since the vectors (a_1, b_1) and (a_2, b_2) are linearly independent, there exists a vector $(c, d) \in \mathbb{R}^2$ satisfying $a_1 c + b_1 d = 1$ and $a_2 c + b_2 d = 0$. Therefore, we obtain that the function $g_1(t) = F(ct, dt) - g_2(0)$ is in the class $C^k(\mathbb{R})$. Similarly, one can verify that $g_2 \in C^k(\mathbb{R})$.

The picture drastically changes if the number of directions $n \geq 3$. For $n = 3$, there are ultimately smooth functions which decompose into sums of very badly behaved ridge

functions. This phenomena comes from the classical Cauchy Functional Equation. This equation,

$$f(x + y) = f(x) + f(y), \quad f : \mathbb{R} \rightarrow \mathbb{R}, \quad (1.4)$$

looks very simple and has a class of simple solutions $f(x) = cx$, $c \in \mathbb{R}$. Nevertheless, it easily follows from the Hamel basis theory that the Cauchy Functional Equation has also a large class of wild solutions. These solutions are called “wild” because they are extremely pathological over reals. They are, for example, not continuous at a point, not monotone at an interval, not bounded at any set of positive measure (see, e.g., [1]). Let g be any wild solution of the equation (1.4). Then the zero function can be represented as

$$0 = g(x) + g(y) - g(x + y). \quad (1.5)$$

Note that the functions involved in (1.5) are bivariate ridge functions with the directions $(1, 0)$, $(0, 1)$ and $(1, 1)$ respectively. This example shows that for smoothness of the representation (1.3) one must impose additional conditions on the representing functions g_i , $i = 1, \dots, n$.

Such additional conditions are recently found by Pinkus [19]. He proved that for a large class of representing functions g_i , the representation is smooth. That is, if apriori assume that in the representation (1.3), the functions g_i belong to a certain class of “well behaved functions”, then they have the same degree of smoothness as the function F . As the mentioned class of “well behaved functions” one may take, e.g., the set of functions that are continuous at a point, bounded on one side on a set of positive measure, monotonic at an interval, Lebesgue measurable, etc. (see [19]). Konyagin and Kuleshov [11] proved that in (1.3) the functions g_i inherit smoothness properties of F (without additional assumptions on g_i) if and only if the directions \mathbf{a}^i are linearly independent. Note that the results of Pinkus and also Konyagin and Kuleshov are valid not only in bivariate but also in multivariate case.

In this paper, we study a different aspect of the problem of representation by ridge functions. Assume in the representation (1.3) $F \in C^k(\mathbb{R}^2)$ but the functions g_i are arbitrary. That is, we allow very badly behaved functions (for example, not continuous at any point). Can we write F as a sum $\sum_{i=1}^n f_i(a_i x + b_i y)$ but with the $f_i \in C^k(\mathbb{R})$, $i = 1, \dots, n$? We see that the answer to this question is positive as expected. For the sake of convenience we state the result over \mathbb{R}^2 , but in fact it holds over any open set in \mathbb{R}^2 .

Note that the above problem is not elementary as it seems. There are cases when representation with good functions is not possible. Such situations happen over closed sets with no interior. In [7], Ismailov and Pinkus presented an example of a function of the form

$$F(x, y) = g_1(a_1 x + b_1 y) + g_2(a_2 x + b_2 y),$$

that is bounded and continuous on the union of two straight lines but such that both g_1 and g_2 are necessarily discontinuous, and thus cannot be replaced with continuous functions f_1 and f_2 .

The result of this paper can be applied to a higher order partial differential equation in two variables if its solution is given by a sum of sufficiently smooth plane waves (see,

for example, Eq. (1.1)). Based on our theorem below, in this case, one can demand only smoothness of the sum and dispense with smoothness of the plane wave summands.

2. Smoothness in bivariate ridge function representation

In this section we prove the following theorem.

Theorem 2.1. *Assume (a_i, b_i) , $i = 1, \dots, n$ are pairwise linearly independent vectors in \mathbb{R}^2 . Assume that a function $F \in C^k(\mathbb{R}^2)$ has the form*

$$F(x, y) = \sum_{i=1}^n g_i(a_i x + b_i y),$$

where g_i are arbitrary univariate functions and $k \geq n - 2$. Then F can be represented also in the form

$$F(x, y) = \sum_{i=1}^n f_i(a_i x + b_i y),$$

where the functions $f_i \in C^k(\mathbb{R})$, $i = 1, \dots, n$.

Proof. Since the vectors (a_{n-1}, b_{n-1}) and (a_n, b_n) are linearly independent, there is a non-singular linear transformation $S : (x, y) \rightarrow (x', y')$ such that $S : (a_{n-1}, b_{n-1}) \rightarrow (1, 0)$ and $S : (a_n, b_n) \rightarrow (0, 1)$. Thus, without loss of generality we may assume that the vectors (a_{n-1}, b_{n-1}) and (a_n, b_n) coincide with the coordinate vectors $e_1 = (1, 0)$ and $e_2 = (0, 1)$ respectively. Therefore, to prove the theorem it is enough to show that if a function $F \in C^k(\mathbb{R}^2)$ is expressed in the form

$$F(x, y) = \sum_{i=1}^{n-2} g_i(a_i x + b_i y) + g_{n-1}(x) + g_n(y), \quad (2.1)$$

with arbitrary g_i , then there exist functions $f_i \in C^k(\mathbb{R})$, $i = 1, \dots, n$, such that F is expressed also in the form

$$F(x, y) = \sum_{i=1}^{n-2} f_i(a_i x + b_i y) + f_{n-1}(x) + f_n(y). \quad (2.2)$$

By $\Delta_l^{(\delta)} f$ we denote the increment of a function f in a direction $l = (l_1, l_2)$. That is,

$$\Delta_l^{(\delta)} f(x, y) = f(x + l_1 \delta, y + l_2 \delta) - f(x, y).$$

We also use the notation $\frac{\partial f}{\partial l}$ which denotes the derivative of f in the direction l .

It is easy to check that the increment of a ridge function $g(ax + by)$ in a direction perpendicular to (a, b) is zero. Let l_1, \dots, l_{n-2} be unit vectors perpendicular to the vectors $(a_1, b_1), \dots, (a_{n-2}, b_{n-2})$ correspondingly. Then for any set of numbers $\delta_1, \dots, \delta_{n-2} \in \mathbb{R}$ we have

$$\Delta_{l_1}^{(\delta_1)} \dots \Delta_{l_{n-2}}^{(\delta_{n-2})} F(x, y) = \Delta_{l_1}^{(\delta_1)} \dots \Delta_{l_{n-2}}^{(\delta_{n-2})} [g_{n-1}(x) + g_n(y)]. \quad (2.3)$$

Denote the left hand side of (2.3) by $S(x, y)$. That is, set

$$S(x, y) \stackrel{\text{def}}{=} \Delta_{l_1}^{(\delta_1)} \dots \Delta_{l_{n-2}}^{(\delta_{n-2})} F(x, y).$$

Then from (2.3) it follows that for any real numbers δ_{n-1} and δ_n ,

$$\Delta_{e_1}^{(\delta_{n-1})} \Delta_{e_2}^{(\delta_n)} S(x, y) = 0,$$

or in expanded form,

$$S(x + \delta_{n-1}, y + \delta_n) - S(x, y + \delta_n) - S(x + \delta_{n-1}, y) + S(0, 0) = 0.$$

Putting in the last equality $x = y = 0$, $\delta_{n-1} = x$, $\delta_n = y$, we obtain that

$$S(x, y) = S(x, 0) + S(0, y) - S(0, 0).$$

This means that

$$\Delta_{l_1}^{(\delta_1)} \dots \Delta_{l_{n-2}}^{(\delta_{n-2})} F(x, y) = \Delta_{l_1}^{(\delta_1)} \dots \Delta_{l_{n-2}}^{(\delta_{n-2})} F(x, 0) + \Delta_{l_1}^{(\delta_1)} \dots \Delta_{l_{n-2}}^{(\delta_{n-2})} F(0, y).$$

By the hypothesis of the theorem, the derivatives $\frac{\partial^{n-2}}{\partial l_1 \dots \partial l_{n-2}} F(x, 0)$ and $\frac{\partial^{n-2}}{\partial l_1 \dots \partial l_{n-2}} F(0, y)$ exist. Denote these derivatives by $h_{1,1}$ and $h_{2,1}$ respectively. Thus, it follows from the above formula that

$$\frac{\partial^{n-2} F}{\partial l_1 \dots \partial l_{n-2}} = h_{1,1}(x) + h_{2,1}(y). \quad (2.4)$$

Note that $h_{1,1}$ and $h_{2,1}$ belong to the class $C^{k-n+2}(\mathbb{R})$.

By $h_{1,2}$ and $h_{2,2}$ denote the antiderivatives of $h_{1,1}$ and $h_{2,1}$ satisfying the condition $h_{1,2}(0) = h_{2,2}(0) = 0$ and multiplied by the numbers $1/\cos(e_1, \wedge l_1)$ and $1/\cos(e_2, \wedge l_1)$ correspondingly. That is,

$$\begin{aligned} h_{1,2}(x) &= \frac{1}{\cos(e_1, \wedge l_1)} \int_0^x h_{1,1}(z) dz; \\ h_{2,2}(y) &= \frac{1}{\cos(e_2, \wedge l_1)} \int_0^y h_{2,1}(z) dz. \end{aligned}$$

Here $(e, \wedge l)$ denotes the angle between vectors e and l . Obviously, the function

$$F_1(x, y) = h_{1,2}(x) + h_{2,2}(y)$$

obeys the equality

$$\frac{\partial F_1}{\partial l_1}(x, y) = h_{1,1}(x) + h_{2,1}(y). \quad (2.5)$$

From (2.4) and (2.5) we obtain that

$$\frac{\partial}{\partial l_1} \left[\frac{\partial^{n-3} F}{\partial l_2 \cdots \partial l_{n-2}} - F_1 \right] = 0.$$

Hence, for some ridge function $\varphi_{1,1}(a_1x + b_1y)$,

$$\frac{\partial^{n-3} F}{\partial l_2 \cdots \partial l_{n-2}}(x, y) = h_{1,2}(x) + h_{2,2}(y) + \varphi_{1,1}(a_1x + b_1y). \quad (2.6)$$

Here all the functions $h_{2,1}, h_{2,2}(y), \varphi_{1,1} \in C^{k-n+3}(\mathbb{R})$.

Set the following functions

$$\begin{aligned} h_{1,3}(x) &= \frac{1}{\cos(e_1, \wedge l_2)} \int_0^x h_{1,2}(z) dz; \\ h_{2,3}(y) &= \frac{1}{\cos(e_2, \wedge l_2)} \int_0^y h_{2,2}(z) dz; \\ \varphi_{1,2}(t) &= \frac{1}{a_1 \cos(e_1, \wedge l_2) + b_1 \cos(e_2, \wedge l_2)} \int_0^t \varphi_{1,1}(z) dz. \end{aligned}$$

Note that the function

$$F_2(x, y) = h_{1,3}(x) + h_{2,3}(y) + \varphi_{1,2}(a_1x + b_1y)$$

obeys the equality

$$\frac{\partial F_2}{\partial l_2}(x, y) = h_{1,2}(x) + h_{2,2}(y) + \varphi_{1,1}(a_1x + b_1y). \quad (2.7)$$

From (2.6) and (2.7) it follows that

$$\frac{\partial}{\partial l_2} \left[\frac{\partial^{n-4} F}{\partial l_3 \cdots \partial l_{n-2}} - F_2 \right] = 0.$$

The last equality means that for some ridge function $\varphi_{2,1}(a_2x + b_2y)$,

$$\frac{\partial^{n-4} F}{\partial l_3 \cdots \partial l_{n-2}}(x, y) = h_{1,3}(x) + h_{2,3}(y) + \varphi_{1,2}(a_1x + b_1y) + \varphi_{2,1}(a_2x + b_2y). \quad (2.8)$$

Here all the functions $h_{1,3}, h_{2,3}, \varphi_{1,2}, \varphi_{2,1} \in C^{k-n+4}(\mathbb{R})$.

Note that in the left hand sides of (2.4), (2.6) and (2.8) we have the mixed directional derivatives of F and the order of these derivatives is decreased by one in each consecutive step. Continuing the above process, until it reaches the function F , we obtain the desired result. \square

Theorem 2.1 can be applied to Eq. (1.1) as follows.

Corollary 2.2. *Assume a function $u \in C^r(\mathbb{R}^2)$ is of the form (1.2) with arbitrarily behaved v_i . Then u is a solution to the Equation (1.1).*

Remark. Some polynomial terms appear while attempting to obtain a smoothness result in multivariate case. In [2], we proved that if a function $f(x_1, \dots, x_n)$ of a certain smoothness class is represented by a sum of r arbitrarily behaved ridge functions, then, under suitable conditions, it can be represented by a sum of ridge functions of the same smoothness class and some n -variable polynomial of a certain degree. The appearance of a polynomial term is mainly related to the fact that in \mathbb{R}^n ($n \geq 3$) there are many directions orthogonal to a given direction. Note that a polynomial term also appears in verifying if a given function of n variables ($n \geq 3$) is a sum of ridge functions (see [4]). However, paralleling the above theorem, we conjecture that if a multivariate function of a certain smoothness class is represented by a sum of arbitrarily behaved ridge functions, then it can also be represented by a sum of ridge functions of the same smoothness class.

References

- [1] J. Aczél, *Functional Equations and their Applications*, Academic Press, New York, 1966.
- [2] R. A. Aliev and V. E. Ismailov, On a smoothness problem in ridge function representation, *Adv. in Appl. Math.* **73** (2016), 154-169.
- [3] E. J. Candès, Ridgelets: estimating with ridge functions, *Ann. Statist.* **31** (2003), 1561-1599.
- [4] P. Diaconis and M. Shahshahani, On nonlinear functions of linear combinations, *SIAM J. Sci. Stat. Comput.* **5** (1984), 175-191.
- [5] J. H. Friedman and W. Stuetzle, Projection pursuit regression, *J. Amer. Statist. Assoc.* **76** (1981), 817-823.
- [6] V. E. Ismailov, Characterization of an extremal sum of ridge functions, *J. Comp. Appl. Math.* **205** (2007), 105-115.
- [7] V. E. Ismailov, A. Pinkus, Interpolation on lines by ridge functions. *J. Approx. Theory* **175** (2013), 91-113.

- [8] V. E. Ismailov, Approximation by ridge functions and neural networks with a bounded number of neurons, *Appl. Anal.* **94** (2015), 2245-2260.
- [9] F. John, *Plane Waves and Spherical Means Applied to Partial Differential Equations*, Interscience, New York, 1955.
- [10] I. Kazantsev, Tomographic reconstruction from arbitrary directions using ridge functions, *Inverse Problems* **14** (1998), 635-645.
- [11] S. V. Konyagin, A. A. Kuleshov, On the continuity of finite sums of ridge functions (Russian), *Mat. Zametki* **98** (2015), 308-309; English transl. in *Math. Notes* **98** (2015), 336-338.
- [12] A. Kroó, On approximation by ridge functions, *Constr. Approx.* **13** (1997), 447-460.
- [13] B. F. Logan and L. A. Shepp, Optimal reconstruction of a function from its projections, *Duke Math. J.* **42** (1975), 645-659.
- [14] V. E. Maiorov, On best approximation by ridge functions, *J. Approx. Theory* **99** (1999), 68-94.
- [15] V. Maiorov and A. Pinkus, Lower bounds for approximation by MLP neural networks, *Neurocomputing* **25** (1999), 81-91.
- [16] F. Natterer, *The Mathematics of Computerized Tomography*, Wiley, New York, 1986.
- [17] P. P. Petrushev, Approximation by ridge functions and neural networks, *SIAM J. Math. Anal.* **30** (1998), 155-189.
- [18] A. Pinkus, *Ridge Functions*, Cambridge Tracts in Mathematics, 205. Cambridge University Press, Cambridge, 2015.
- [19] A. Pinkus, Smoothness and uniqueness in ridge function representation, *Indag. Math. (N.S.)* **24** (2013), no. 4, 725-738.
- [20] A. Pinkus, Approximation theory of the MLP model in neural networks, *Acta Numerica* **8** (1999), 143-195.